

Análisis de co-términos y de redes sociales para la generación de mapas temáticos

Por Sandra Miguel, Lorena Caprile e Israel Jorquera-Vidal



Sandra Miguel, licenciada en Bibliotecología y Documentación, posee un diploma de Estudios Avanzados de Tercer Ciclo del Doctorado en Documentación e Información Científica, Universidad de Granada, España. Es profesora de la Licenciatura en Bibliotecología y Ciencia de la Información, Universidad Nacional de La Plata, Argentina.



Lorena Caprile es bibliotecaria-documentalista en la Biblioteca de la Facultad de Ciencias Naturales y Museo de la Universidad Nacional de La Plata, Argentina. Imparte clases de iniciación a la docencia e investigación en la Licenciatura en Bibliotecología y Ciencia de la Información, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata.



Israel Jorquera-Vidal es bibliotecario-documentalista en la Biblioteca de la Facultad de Ciencias Naturales y Museo de la Universidad Nacional de La Plata. Especialista en diseño y desarrollo de sitios web.

Resumen: La mejora de las capacidades de búsqueda y de las interfaces de los opacs sigue siendo uno de los principales desafíos para las bibliotecas, especialmente en lo que respecta al acceso por materias. Las interfaces visuales pueden facilitar la recuperación. El objetivo del presente trabajo es explorar si la combinación de técnicas de análisis de co-términos y de redes sociales resulta ser una metodología válida para la generación de mapas temáticos de la colección. La principal conclusión es que el método es válido, y que los mapas obtenidos podrían servir como interfaz visual para el acceso por materias. También es útil para detectar problemas en los criterios de indización y contribuir a la mejora de la calidad de la descripción temática del conjunto documental.

Palabras clave: Análisis de redes sociales, Análisis de co-términos, Opacs, Visualización de información, Interfaces visuales, Recuperación de información.

Title: Co-term and social networks analysis for the generation of subject maps

Abstract: Improvement of the interfaces and search capacities of the opacs continues to be one of the main challenges for libraries, especially with regard to searching by subject. The visual interfaces can help with retrieval. This paper explores the combination of co-term analysis techniques with social networks for the generation of subject maps of collections. The main conclusion is that the methodology applied is valid, and that the maps could be used for browsing and subject retrieval. It would also be useful for improving the quality of indexing.

Keywords: Social network analysis, Co-term analysis, Opacs, Information visualization, Visual interfaces, Information retrieval.

Miguel, Sandra; Caprile, Lorena; Jorquera-Vidal, Israel. "Análisis de co-términos y de redes sociales para la generación de mapas temáticos". En: *El profesional de la información*, 2008, noviembre-diciembre, v. 17, n. 6, pp. 637-646.

DOI: 10.3145/epi.2008.nov.06

1. Introducción

Las investigaciones sobre opacs (*online public access catalog*) durante los últimos veinte años revelan una incesante búsqueda por encontrar las mejores formas de representar la información, ya que actúan de intermediarios entre los usuarios y la bibliografía (White; McCain, 1997). Los principales desafíos están vinculados a la mejora de las capacidades de búsqueda

de los sistemas de recuperación de información, y al diseño de sus interfaces (Mi, 2008).

A ello se suman antiguos problemas aún no resueltos, como los relacionados con la búsqueda por materias (Drabenstott; Weller, 1996). Conforme a los resultados de algunos estudios, este tipo de búsqueda es la más solicitada por los usuarios, y paradójicamente es la que mayores dificultades plantea (Larson, 1991;

Artículo recibido el 14-08-08

Aceptación definitiva: 14-10-08

Ríos-García, 1991; Su, 1994; Pichinini y Martínez, 1996). En el 40% de los fracasos en las búsquedas se debe a la sobrecarga o exceso de resultados; o por el contrario, a resultados nulos por la falta de coincidencia entre los términos empleados en la recuperación, y los almacenados en el sistema. Ante este problema, un importante porcentaje de usuarios ha expresado su deseo de ver las palabras relacionadas con su tema, y así poder elegir las para su expresión de búsqueda.

“La mejora de las interfaces de los opacs sigue siendo un desafío para las bibliotecas”

Ello nos lleva a pensar que una interfaz que ofrezca al usuario la posibilidad de visualizar las materias del opac, podría ser una buena solución para reducir el fracaso en las búsquedas. El objetivo de este trabajo es explorar si la combinación de técnicas de análisis de co-términos y de redes sociales es válida para la generación de mapas temáticos de las colecciones de las bibliotecas, y presentar sus principales ventajas y limitaciones.

La interfaz usuario-sistema es el lenguaje de comunicación entre el hombre y la máquina (Matthews, 1983). Su propósito es facilitar la tarea de recuperación de información, reduciendo los errores humanos y mejorando el desempeño del usuario durante el proceso de la búsqueda de información.

De acuerdo con la clasificación propuesta por Hildreth (1985), la segunda generación de opacs incorporó muchas funcionalidades y elementos de diseño de la interfaz orientados a mejorar las prestaciones de búsqueda por materia. Algunos ejemplos son la posibilidad de realizar consultas por campos y la inclusión de índices de materias. Se ha demostrado que, cuando estos índices tienen un mantenimiento regular y el usuario conoce bien su manejo, constituyen una herramienta de gran utilidad para obtener resultados exitosos (Villén, 2006).

“Combinando técnicas de análisis de co-términos y de redes sociales se pueden generar mapas temáticos”

A pesar de estos avances, algunos autores sostienen que para conseguir mayor usabilidad las interfaces de

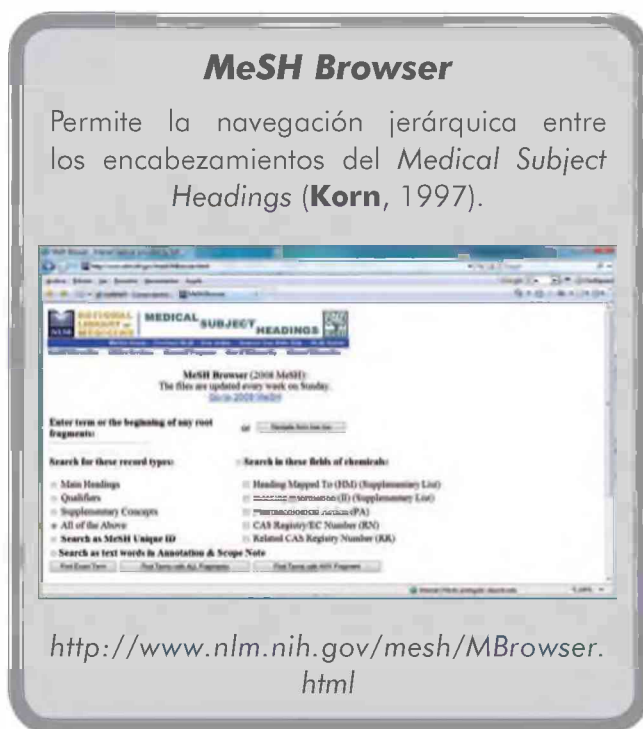
los opacs aún deben dar un salto cualitativo y cuantitativo (Játiva, 2004).

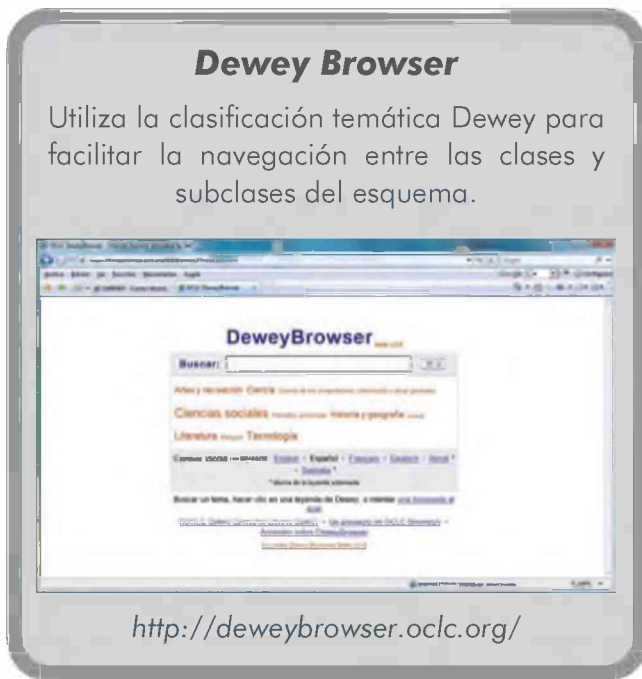
La manera que se elija para presentar al usuario final la información es tan importante como el contenido en la recuperación de información. La forma en la que se haga accesible en pantalla determinará que la información cumpla su finalidad de ser útil, o por el contrario se quede almacenada sin más (Marcos, 2003).

Particularmente cuando el usuario no tiene claro lo que está buscando, o cuando tiene dificultades para formalizar su necesidad de información a través del lenguaje de interrogación del sistema, se requiere de un modelo alternativo o complementario que le brinde otra vía de acceso al conocimiento. Este modelo alternativo, que puede ser una exploración o navegación (*browsing*), da al usuario la posibilidad de inspeccionar el conjunto documental sin tener que expresar de forma previa sus necesidades con una o más palabras (Herreo-Solana; Hassan-Montero, 2006).

En la mayoría de los opacs actuales las materias que representan temáticamente la colección se presentan al usuario como un listado de términos dispuestos en orden alfabético. En el mejor de los casos se han incorporado ficheros de autoridades o tesauros con presentaciones jerárquicas.

La superación de estas estrategias de exploración viene de la mano de las investigaciones sobre Interfaces visuales de recuperación de información (Viri). El principal objetivo de las mismas es la representación visual e interactiva de la información en un espacio de dimensiones reducidas, resultado de la explotación de





los datos (*data mining*) y de la abstracción gráfica del conjunto documental.

Para Noyons y otros (1999) el uso de mapas como metáforas de un campo de conocimiento tiene un gran potencial como interfaz para la recuperación de la información. Por su parte, Janecek (2005) plantea que las interfaces visuales son potencialmente útiles para que los usuarios puedan explorar una representación de la colección y oportunamente descubrir información que les guíe a encontrar los documentos para ellos relevantes.

Desde esta perspectiva, la representación visual de los esquemas de clasificación temática permite a los usuarios tener un mapa mental de cómo está organizada la biblioteca, lo que suponemos que facilitará el acceso por materias.

La visualización de información no es el resultado inmediato del acto de “ver”. Requiere transformar datos abstractos y fenómenos complejos de la realidad en mensajes visibles. Esto hace posible que los individuos vean con sus propios ojos datos y fenómenos que yacen ocultos y que no son directamente aprehensibles. En otras palabras, la visualización consiste en hacer visible para nuestra mente aquello que no es visible para nuestros ojos (Vargas-Quesada; Moya-Anegón, 2007).

“Una interfaz donde visualizar las materias probablemente ayudaría a reducir el fracaso en las búsquedas”

2. Interfaces visuales en recuperación de información

En los últimos años varios autores han estudiado las estrategias e iniciativas que desarrollan interfaces visuales. Moya-Anegón y Herrero-Solana (1999) analizaron proyectos de investigación sobre interfaces gráficas bidimensionales y tridimensionales que permitieran la exploración de grandes volúmenes de información. Marcos (2004) revisó y describió con detalle otros proyectos, diferenciándolos según estuvieran orientados a la interfaz de acceso o a la forma de presentación de los resultados. Hassan-Montero y Herrero-Solana (2006) analizaron diferentes estrategias para elaborar interfaces visuales, todas ellas basadas en la identificación de las relaciones semánticas de similitud existentes entre documentos, o incluso entre términos. Asimismo, White y McCain (1997) y Börner y otros (2003) escribieron dos excelentes revisiones sobre técnicas de visualización. También nosotros, en un intento por mostrar la diversidad metodológica de las diferentes propuestas, mencionaremos algunas de las iniciativas.

Entre ellas están los mapas auto-organizativos (*self-organizing maps* o *SOM*). Son mapas basados en un tipo de red neuronal que trabaja bajo el principio de auto-organización y clasificación automática de la información. Cada nodo en el mapa está asociado a una lista de documentos semánticamente similares, y la proximidad espacial entre ellos indica una similitud conceptual (Börner, 2003). Las primeras aplicaciones al campo de la recuperación de información se atribuyen a Xia Lin y otros (Lin; Soergel y Marchionini, 1991); los prototipos más conocidos son *WebSom*, dirigido por Kohonen y *ET-Map*, dirigido por Chen. Sin embargo no hemos encontrado un sistema que utilice estos mapas como interfaz de recuperación. Algunos autores han señalado que su construcción exige una gran capacidad de procesamiento y que presenta limitaciones en cuanto a la cantidad de información que puede procesar (Guerrero-Bote, 2002).

<http://websom.hut.fi/>

Otros proyectos apuestan por la generación de interfaces utilizando técnicas de análisis de cluster (*clustering*), las cuales permiten agrupar documentos en función de patrones de proximidad o distancia, obteniendo como resultado clases o clusters de documentos similares. *Grokker* es un buscador web cuya interfaz visual permite al usuario navegar entre los clusters. A pesar de que es visualmente atractivo, tiene como desventaja que a medida que se va descendiendo en niveles de agregación más específicos, se va perdiendo la vista del todo, aspecto que a nuestro juicio produce cierta desorientación.

“Las representaciones gráficas de las redes sociales son ideales como metáfora visual para la recuperación de información”

Otras propuestas de modelos de interfaces surgen de la combinación de varios métodos. Una combina técnicas de minería de metadatos, de escalamiento multidimensional y una técnica de *focus+context* denominada “ojo de pez” (Hassan-Montero y Herreiro-Solana, 2006). Su interés radica en que los autores llevaron a cabo un prototipo y una prueba con usuarios reales, quienes valoraron de forma positiva la interfaz visual propuesta.

Otro tipo de representación visual que ha cobrado especial énfasis en los últimos años es la basada en la teoría de grafos.

Un grafo es un conjunto de nodos con unas relaciones explícitas entre ellos (Caraballo-Pérez, 2007). Los nodos representan actores (documentos, autores, palabras, términos, etc.), y los enlaces relaciones de algún tipo.

TopicMaps, por ejemplo, se basa en ellos para estructurar la información y presentarla en la forma de una red semántica. Una interfaz que lo implementa es *Aquabrowser*, un sistema que permite visualizar, rápida y organizadamente, un conjunto de términos relacionados conceptualmente con los propuestos por el usuario (Morillo-Velarde, 2006; Senso, 2006). Lo interesante es que ya no es un prototipo, sino que está en funcionamiento en muchas bibliotecas, como *Queens Library* o *Arlington Public Library*

<http://aqua.queenslibrary.org/>

<http://www.acornweb.org/>

Mapa semántico de una colección

Pero *Aquabrowser* no permite aún obtener una representación sintética de la colección de la biblioteca. Hace ya más de diez años Schneiderman (1996) señalaba que las interfaces de los sistemas de recuperación de información deberían proporcionar una primera representación gráfica o mapa, que sea lo suficientemente general e informativo para que el usuario pueda tener una idea global del dominio de conocimiento en cuestión, para luego ir haciendo *zoom* y descender a niveles cada vez más específicos.

El *Atlas de la ciencia*, del grupo español *SCImago*, utiliza una interfaz visual que aplica esta estrategia para la representación de dominios científicos. Utiliza un conjunto de mapas como interfaz para que el usuario navegue entre categorías temáticas, en distintos niveles

de agregación, hasta llegar a los documentos propiamente dichos (Moya-Anegón y otros, 2004). En este caso, la combinación de técnicas de análisis de co-citación y de redes sociales ha demostrado ser una estrategia válida para la generación automática de estas interfaces visuales e interactivas.

<http://www.atlasofscience.net/>

3. Análisis de co-términos

Por co-términos entendemos la co-ocurrencia o aparición conjunta de dos términos en un texto dado. El análisis de co-términos es una variante del análisis de co-palabras (un término puede estar compuesto de varias palabras). En cualquier caso, se trata de una técnica de análisis de contenido que usa patrones de co-ocurrencia de pares de ítems (términos o palabras) dentro de un corpus textual, para identificar las relaciones entre las ideas en un dominio determinado. En nuestro caso el corpus es cada registro bibliográfico del opac, y los términos las expresiones utilizadas para la descripción del contenido de los documentos, y que son incluidos en los campos de materia de la base de datos. Desde esta perspectiva de análisis, dos términos (y por extensión dos documentos) estarán más vinculados o asociados conceptualmente entre sí, cuanto mayor sea la co-ocurrencia entre ellos.

La técnica de análisis de co-palabras fue desarrollada por Michel Callon y Bruno Latour (Callon, 1986; Latour, 1992). Aunque fue primeramente propuesta con la finalidad de mapear la dinámica de los campos científicos, el propio Callon señaló su utilidad como un método para clasificar documentos con fines de recuperación de información, y construcción y actualización de tesauros (He, 1999). Más recientemente algunos autores propusieron su uso para la creación de interfaces de bibliotecas digitales (Buzydlowski y otros, 2002); (Börner y Chen, 2002).

4. Análisis de redes sociales

El análisis de redes sociales es una nueva perspectiva conceptual y metodológica del campo de las ciencias sociales y del comportamiento. Analiza las relaciones existentes entre los elementos, a diferencia del enfoque tradicional de análisis atributivo de los elementos de un grupo. Los datos son relacionales, entendiendo por dato relacional un vínculo específico existente entre un par de elementos. Desde esta perspectiva, la relación entre pares es una propiedad del par (es decir de la relación) y no una cuestión inherente a las características individuales de cada uno de los elementos relacionados (Wasserman y Faust, 1998).

Una red social está conformada por al menos dos componentes: los actores, que son las unidades que se

relacionan entre sí por algún motivo, finalidad o circunstancia, y las relaciones, que son los lazos o vínculos existentes entre los actores.

Los actores de una red pueden desempeñar diferentes roles según sea su grado, poder de intermediación en relación con los demás, cercanía, etc. Asimismo, los vínculos entre ellos pueden ser directos o indirectos, direccionales o no-direccionales, tener diferentes intensidades; aspectos todos que van determinando el tipo de relación existente y el tipo de estructura de red que conforman (Rodríguez, 1995). A cada actor de la red se le identifica con un símbolo (generalmente un círculo), y su posición en la estructura está determinada por el rol que desempeña.

Las representaciones gráficas de las redes sociales son grafos. Como son relativamente fáciles de comprender, son ideales como metáfora visual para la representación de información (Dürsteler, 2003), además de que permiten descubrir relaciones que de otra forma permanecerían ocultas (Vargas-Quesada, 2005).

La creación de conocimiento es algo social, donde un actor puede ser alguien o algo creador de asociaciones (un científico, un texto o un concepto) y las asociaciones o relaciones no son estáticas sino que cambian y se transforman constantemente (Buchelli-Guerrero, 2008). Desde estos supuestos, una red social de términos de un campo temático podría representar la estructura de conocimiento de dicho campo.

Si trasladamos estas ideas a las bibliotecas, los términos que describen temáticamente los documentos de una colección conforman un entramado o red cuya evolución está determinada por la dinámica de su desarrollo.

“Las representaciones visuales de los esquemas de clasificación temática permiten a los usuarios tener un mapa mental de cómo está organizada la biblioteca”

5. Aplicación de las técnicas

Para determinar si la combinación de técnicas resulta válida para nuestro objetivo hemos utilizando como fuentes de datos dos bibliotecas de facultad de la *Universidad Nacional de La Plata*, Argentina. Elegimos estas instituciones porque sus colecciones representan temáticas de campos disciplinares con patrones de producción de conocimientos diferentes. Una está especializada en ciencias naturales (*Cnat*) y la otra en ciencias humanas y sociales (*Hums*), lo que permite

evaluar si la metodología es aplicable a colecciones de temáticas distintas.

En total hemos analizado las colecciones de cuatro catálogos:

- *LCnat* (libros de ciencias naturales);
- *LHums* (libros de humanidades y ciencias sociales);
- *RCnat* (revistas de ciencias naturales) y
- *RHums* (revistas de humanidades y ciencias sociales).

La población estudiada está constituida por todos los registros de los catálogos de revistas (5.249 para *RCnat* y 3.002 para *RHums*), y el subconjunto de registros de los catálogos de libros que tenían completos los campos de materia (48% para *LCnat* -9.350 de un total de 19.623 registros- y 45% para *LHums*, -29.115 de un total de 64.822-). De cada uno de los registros extrajimos los términos que representan los temas de los documentos en dos niveles de agregación, uno más general y otro más específico. El primero se identifica con el nivel de disciplina (geología, biología, antropología, historia, literatura, etc.), y el segundo con palabras clave de naturaleza y grado de especificidad diversa. Cabe señalar que *Hums* utiliza un tesoro para la indización, y *Cnat* una lista de términos elaborada *ad hoc*.

En ambos casos hubo que realizar un proceso de normalización de las entradas para calcular las frecuencias de ocurrencia de cada término. Seleccionamos el subconjunto de aquellos con mayor presencia, debido a que una red con una cantidad excesiva de nodos y relaciones presenta serias dificultades de interpretación. Determinar dónde establecer el punto de corte no fue tarea sencilla, si bien hay varios elementos a considerar. Por una parte, que la cantidad de nodos no sea excesivamente grande para facilitar la clara identificación de cada uno de ellos. Por otra, que no es posible excluir nodos que actúan como puente de otros nodos. Éstos reciben el nombre de puntos de corte o intermediarios, justamente porque quitarlos produce la desconexión de otros nodos de la estructura, y un incremento del número de componentes de la red. Considerando estos aspectos fijamos en cada caso un umbral donde no sólo se considere el peso individual de cada término, sino su relación con el resto de los términos que conforman la estructura.

Una vez seleccionados los términos construimos una matriz cuadrática y simétrica con las frecuencias de co-ocurrencia de cada uno con el resto, y generamos el mapa temático de cada colección analizada.

Para los conteos de co-ocurrencia aplicamos como criterio que por cada término de nivel más general (nivel de disciplina) aparezca siempre asociado uno de

nivel más específico (palabra clave), siempre que estuvieran presentes en el mismo registro bibliográfico. Por ejemplo, si un registro del opac tiene asignado como término genérico *antropología*, y como palabras clave *etnología* y *folklore*, las asociaciones que se establecen son: *antropología* <> *etnología*; *antropología* <> *folklore*; y *folklore* <> *etnología*. Con ello pretendemos que la red no sólo refleje las relaciones entre los términos más específicos, sino que también permita visualizar las jerarquías que se establecen entre los términos. Las jerarquías responden al típico esquema de organización del conocimiento de las colecciones de las bibliotecas, basado en la *Clasificación Decimal Universal (CDU)*, la *Clasificación Dewey (CDD)* u otros semejantes.

La construcción de las redes y las representaciones gráficas las realizamos utilizando el programa *Pajek*. Para una mejor visualización de la estructura de red empleamos el algoritmo *Kamada-Kawai (Kamada; Kawai, 1989)*, y a efectos de facilitar la interpretación y hacer más legibles las relaciones entre los nodos (términos) eliminamos los enlaces (líneas) más débiles.

El tamaño de los nodos representa el peso que cada término tiene en la red, y el color y grosor de las líneas un indicador de la intensidad de la relación.

6. Análisis e interpretación de los resultados

Como resultados del análisis obtuvimos cuatro mapas temáticos de las colecciones de las bibliotecas (Figuras 1, 2, 3 y 4). Los mapas presentan tanto rasgos comunes como diferenciados.

Entre los comunes podemos mencionar:

Por un lado, que los nodos (términos) que más vinculaciones tienen con el resto tienden a ocupar posi-

ciones centrales en el mapa, mientras que aquellos con menos relaciones se sitúan en la periferia. Este hecho responde a una característica de las representaciones obtenidas a partir de análisis de redes sociales, en virtud de que desde esta perspectiva de análisis los actores o nodos centrales en la estructura no son necesariamente aquellos que tienen mayor presencia individual (en nuestro caso más alta frecuencia de ocurrencia), sino aquellos que tienen más vinculaciones con los demás (en nuestro caso, co-ocurrencia con más cantidad de términos).

La disposición de los nodos en cada red guarda relación estrecha con la cantidad e intensidad de las asociaciones que cada uno de ellos tiene con el resto.

Por otro lado, los nodos tienden a agruparse en racimos o grupos disciplinares. Los términos que representan a las disciplinas desempeñan un importante rol en la red, actuando como núcleos temáticos alrededor de los cuales orbitan los términos más específicos relacionados más intensamente con ellos. Asimismo, en cada red vemos que hay algunos grupos temáticos que tienen mayor presencia que otros, y esos son precisamente los que revelan cuál es el perfil temático de la colección. Para facilitar la rápida identificación de los términos que cumplen esa función los hemos resaltado con negrita.

Otra característica común a todos los mapas es el tipo de estructura. En todos es más jerárquica que reticular. Salvo contadas excepciones los términos específicos se vinculan bastante poco entre ellos. Ello evidencia sin duda el criterio empleado por las bibliotecas para organizar temáticamente sus colecciones, basado en esquemas jerárquicos del tipo *CDU*, *CDD*, u otros similares. Que este tipo de esquema se vea reflejado en los mapas no es casual, sino que

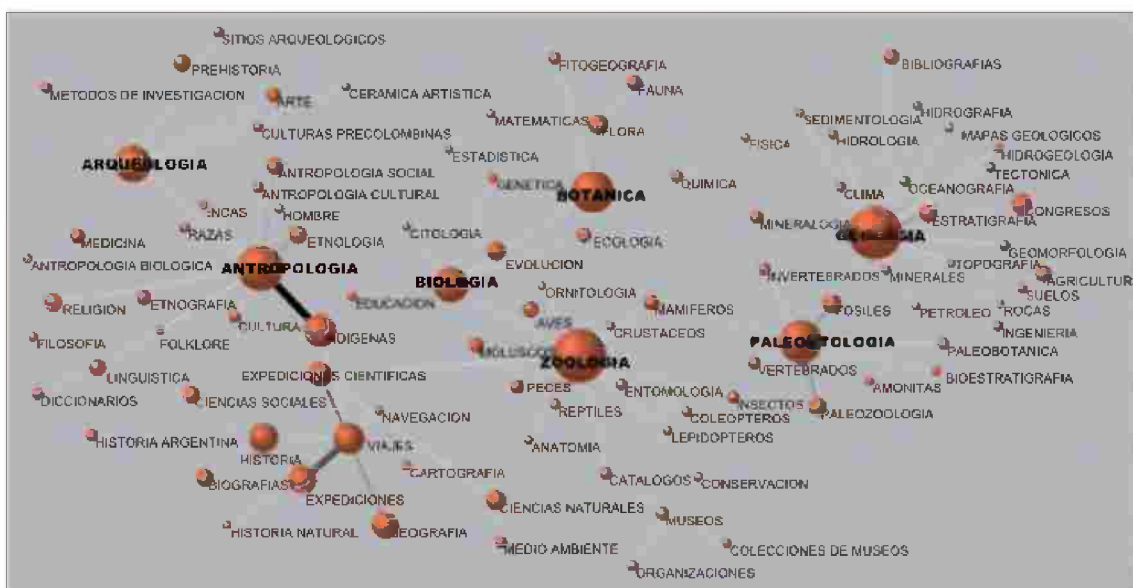


Figura 1. Mapa temático de la colección de libros de ciencias naturales – LCnat

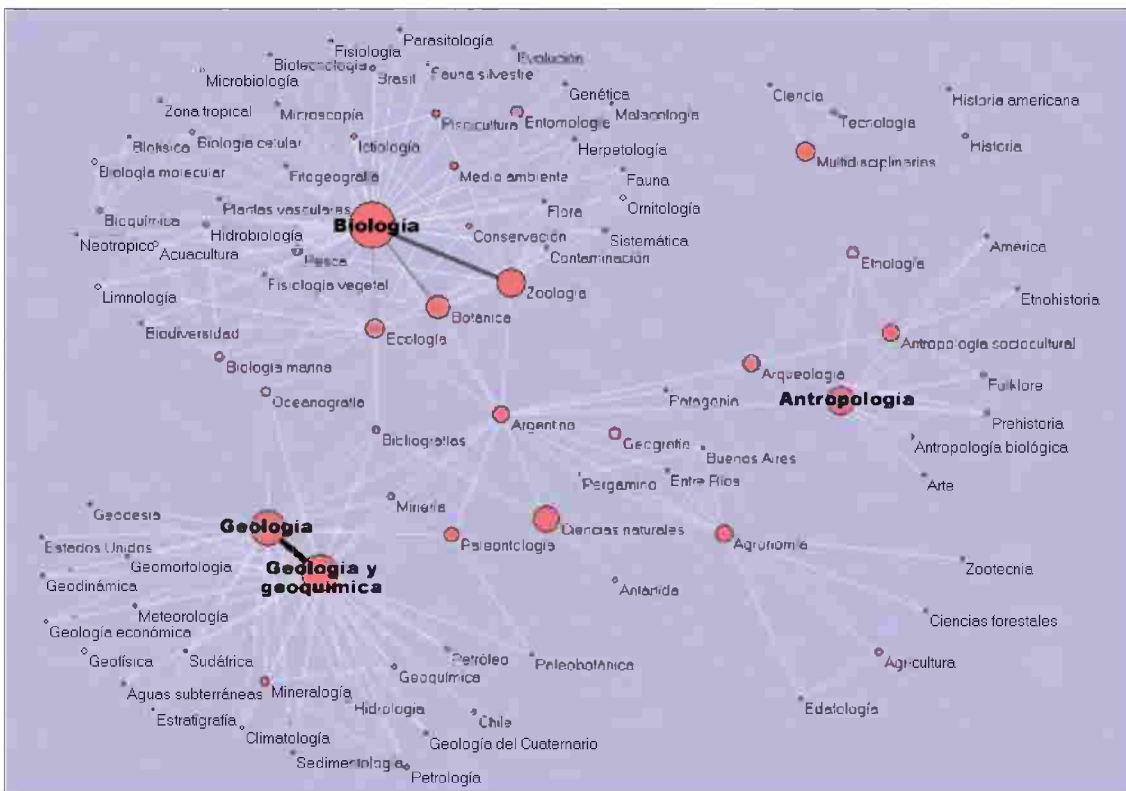


Figura 2. Mapa temático de la colección de revistas de ciencias naturales – RCnat

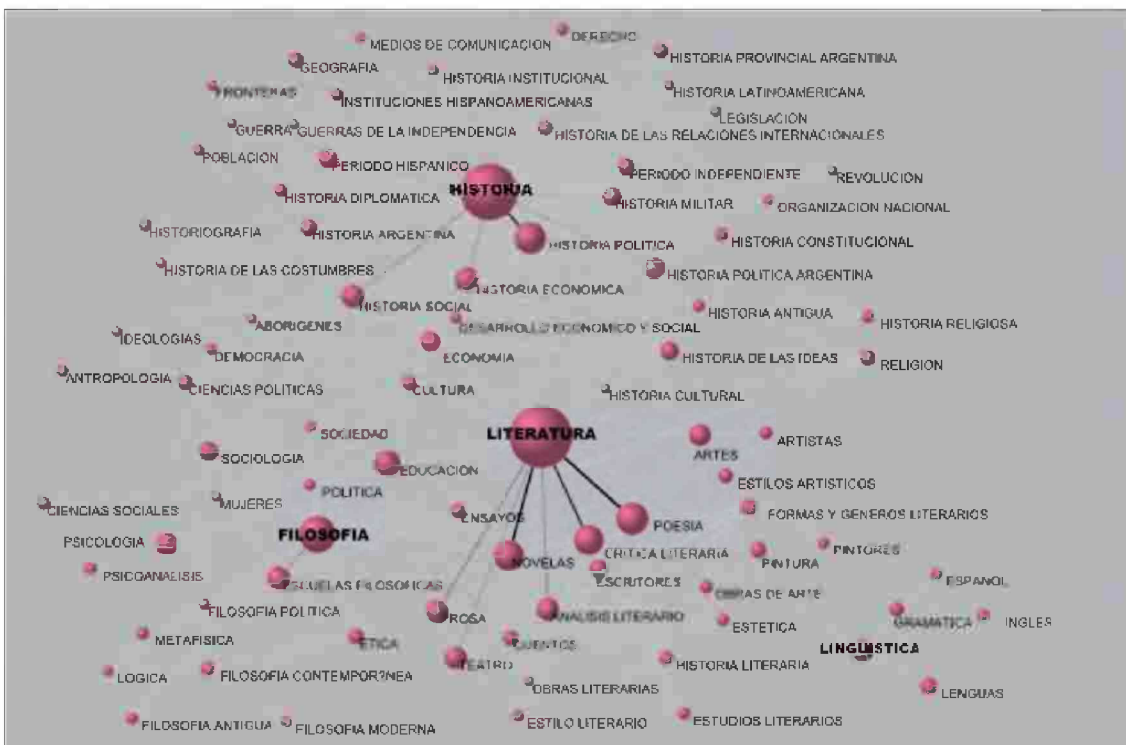


Figura 3. Mapa temático de la colección de libros de humanidades y ciencias sociales – LHums

responde a la metodología que hemos empleado para construirlos, basada en la asociación de un término genérico por cada uno de los términos específicos utilizados para la indización. Asimismo, otro factor asociado a este fenómeno es la cantidad de términos asignados a cada registro. En el caso de los catálogos analizados encontramos que *RHums* tenía un prome-

dio de 2,4 términos por registro; *LHums* 5,3; *LCnat* 3,5; y *RCnat* 2,5. Es evidente que las revistas son más pobres en su descripción temática, y que cuantos más términos se usen para la indización, mayor será la probabilidad de establecer asociaciones entre ellos. No por casualidad la red de mayor densidad es la de *LHums*.

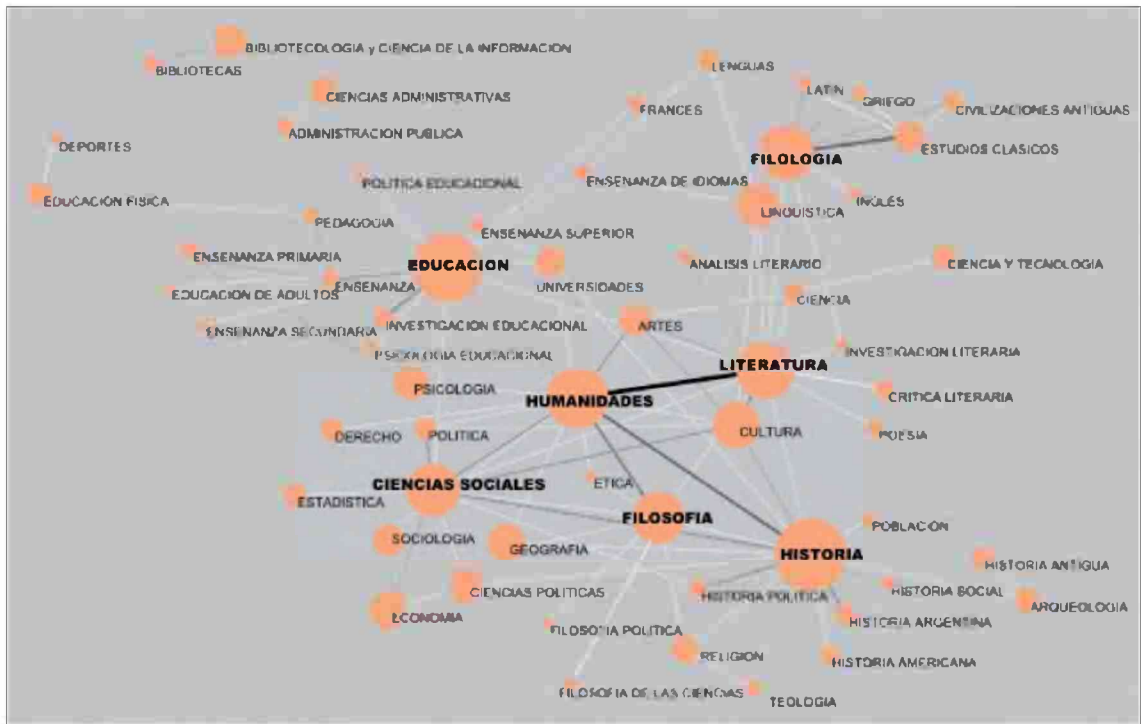


Figura 4. Mapa temático de la colección de revistas de humanidades y ciencias sociales – RHums

La dificultad de encontrar relaciones de co-ocurrencia en bases de datos referenciales no es un tema menor para poder aplicar estas técnicas con fines de visualización, y podría explicar el hecho de que la mayoría de los prototipos de interfaces visuales de sistemas de recuperación de información se hayan planteado para bibliotecas digitales (Marcos, 2004).

En cuanto a las principales diferencias encontradas en los mapas podemos mencionar las siguientes:

En primer lugar destaca que en el caso de *Cnat* se detecta que los criterios de indización empleados no son los mismos. La diferencia más notoria se puede visualizar en el caso de la biología, que en *RCnat* (figura 2) nuclea a la zoología y a la botánica, y en *LCnat* (figura 1) cada uno aparece como un nodo central de un grupo temático bien delimitado.

Para encontrar una explicación a ese fenómeno consultamos a los catalogadores de la biblioteca y extrajimos una serie de conclusiones. En ambos casos utilizan una lista de términos *ad hoc*, pero en *RCnat* la indización había sido cuidadosamente revisada: normalizando términos, reduciendo la cantidad de entradas diferentes, y respetando siempre un mismo criterio de asignación de descriptores que involucraba a la biología siempre como término tope de los otros (zoología y botánica). En el caso de *LCnat*, sin embargo, la indización se hacía de manera más libre y no se tenía mucho conocimiento sobre cuán coherente y consistente era el criterio utilizado para llevarla a cabo.

En los catálogos de *Hums* parece respetarse un mismo criterio desde la perspectiva de la elección de los

términos, y ello posiblemente se deba a que la indización en esta biblioteca se realiza utilizando un tesaurus. No obstante, las diferencias radican en el promedio de términos asignados: en *RHums* 2,4 mientras que en *LHums* es de 5,3.

Las distinciones observadas no parecen triviales a la hora de pensar en estos mapas como posibles interfaces de búsqueda. Sin embargo, resulta evidente que el problema está en la indización y no en el método utilizado para la representación visual de la red temática de la colección. Es obvio que la calidad de los resultados de este tipo de análisis depende de una variedad de factores. Algunos de ellos son la cobertura de la base de datos y la adecuación de los métodos estadísticos escogidos. Pero el principal factor es la cantidad de términos y la coherencia de la indización (He, 1999), sin lo cual difícilmente puede obtenerse una buena representación visual de la colección.

Por el contrario, se han observado limitaciones inherentes al método. Por ejemplo que para obtener mapas legibles y fáciles de interpretar resulte necesario acotar la cantidad de nodos y enlaces de las redes. Ello significa que no podemos representar la totalidad de términos que describen el fondo documental, sino sólo aquellos con mayor presencia o con mayor cantidad de lazos.

Validación del método

Ambos criterios son metodológicamente válidos. En primer lugar, la selección de términos por sus propiedades atributivas parte del supuesto de que sólo los nodos con más altas frecuencias de aparición integra-

rán la red. El segundo criterio circunscribe la población objeto de estudio a los actores que tienen relaciones entre sí (Miceli, 2008). En nuestro caso hemos considerado ambos criterios. Seleccionamos los términos más frecuentes procurando no excluir los que poseen un alto poder de intermediación.

El recorte en la cantidad de términos también hubiera tenido lugar si en vez de análisis de redes sociales hubiésemos empleado técnicas de clustering o escalamiento multidimensional. Como hemos comentado previamente, el principal problema radica, en todo caso, en la determinación del punto de corte, que no sólo tiene que ver con las frecuencias de ocurrencia sino con el rol que desempeña cada nodo en la red. En este sentido, un claro ejemplo de nodos con una gran capacidad de intermediación son los topónimos que aparecen en el mapa de la figura 2. El ejemplo más significativo es el término Argentina, nodo cuya exclusión desvirtuaría considerablemente la estructura temática de la colección de *RCnat*. En los otros casos, los nombres geográficos no tenían esa función como tampoco altas frecuencias de ocurrencia, y por tanto quedaron excluidos de los mapas.

Cabe señalar asimismo que aunque para otros fines los términos con altas frecuencias de ocurrencia suelen ser excluidos por su escaso poder de discriminación (Peat; Willett, 1991), para los objetivos de este estudio dichos términos cumplen una función esencial, en tanto nodos que demarcan los grupos temáticos más representativos de la colección.

7. Conclusiones

Las principales conclusiones a las que llegamos con este estudio son:

Que la combinación de las técnicas de análisis de co-términos y de redes sociales resulta ser una alternativa metodológica válida para la creación de mapas temáticos de las colecciones de las bibliotecas, al menos de bibliotecas universitarias.

Que la metodología resultó ser apropiada para representar colecciones de tamaños, tipos documentales y temáticas diferentes.

Que es posible generar mapas temáticos de las colecciones, independientemente de que las bibliotecas utilicen o no un tesoro para la indización.

Asimismo, fue posible comprobar que aún con una cantidad reducida de términos por registro, es posible generar una representación gráfica de la estructura temática de la colección. No obstante, también se evidencia que cuantos más términos se asignen a un documento, mayores serán las posibilidades de establecer asociaciones temáticas entre ellos.

Que aunque no sea posible visualizar en un mapa la totalidad de los términos que describen el fondo documental, el método es viable para reflejar la estructura temática nuclear de la colección y pone de relieve los grupos disciplinares más representativos. De igual modo, los mapas revelan que hay que seguir perfeccionando la metodología para mejorar la legibilidad.

Por último, creemos que este tipo de mapas podría servir tanto como una alternativa de interfaz visual para el acceso por materias a las colecciones de las bibliotecas, como también para detectar problemas en los criterios de indización, y contribuir a la mejora de la calidad de la descripción temática del conjunto documental. Resta ahora poner a prueba si esta estrategia de representación también resulta válida y útil para el usuario final.

8. Bibliografía

- Börner, Katy; Chen, Chaomei; Boyack, Kevin W.** "Visualizing knowledge domains". *Annual review of information science and technology*, 2003, v. 37, pp. 179-255.
- Börner, Katy; Chen, Chaomei.** "Visual interfaces to digital libraries. Motivation, utilization, and socio-technical challenges". *Lecture notes in computer science*. 2002, n. 2539, pp. 1-9.
- Buchelli-Guerrero, Víctor-Andrés.** *Borrado Cero – Herramienta informática de vigilancia tecnológica para el análisis de comunidades científicas*. Maestría en Ingeniería de Sistemas y Computación. Universidad Nacional de Colombia, 2006. Consultado en: junio 2008. <http://dis.unal.edu.co/profesores/ypinzon/2013326-206/docs/Tesis0Bucheli.pdf>
- Buzydowski, Jan W.; White, Howard D.; Lin, Xia.** "Term co-occurrence analysis as an interface for digital libraries". *Lecture notes in computer science series*, 2002, v. 2539, pp. 133-144.
- Callon, Michel.** "Éléments pour une sociologie de la traduction, la domestication des coquilles Saint-Jacques et des martins-pêcheurs dans la baie de Saint-Brieuc". *L'année sociologique*, 1986, v. 36, pp. 169-208.
- Caraballo-Pérez, Yeter.** "Los topic maps y su relación con las redes sociales". *Acimed*, 2007, v. 16, n. 4. Consultado en: junio 2008. http://www.bvs.sld.cu/revistas/aci/vol16_4_07/aci091007.html
- Chen, Chaomei.** "Bridging the gap: The use of Pathfinder networks in visual navigation". *Journal of visual languages and computing*, 1998, v. 9, n. 3, pp. 267-286. Consultado en: agosto 2008. <http://www.pages.drexel.edu/~cc345/papers/fvlc.pdf>
- Drabenstott, Karen; Weller, Marjorie S.** "Failure analysis of subject searches in a test of a new design for subject access to online catalogs". *Journal of the American Society for Information Science*, 1996, v. 47, n. 7, pp. 520-537.
- Dürsteler, Juan C.** "Visualización de redes sociales". *InfoVis.net*, 2003, n. 136. Consultado en: junio 2008. <http://www.infovis.net/printMag.php?num=136&lang=1>
- Guerrero-Bote, Vicente-P.; Moya-Anegón, Félix de; Herrero-Solana, Víctor.** "Document organization using Kohonen's algorithm". *Information processing and management*, 2002, v. 38, pp. 79-89.
- Hassan-Montero, Yusef; Herrero-Solana, Víctor.** "Interfaz visual para recuperación de información basada en análisis de metadatos, escalamiento multidimensional y efecto ojo de pez". *El profesional de la información*, 2006, v. 15, n. 4, pp. 278-289. Consultado en: abril 2008. http://eprints.relis.org/archive/00009547/01/vol15_4.4.pdf
- Hassan-Montero, Yusef.** "Visualización y recuperación de información". *II Encontro de Ciências e Tecnologias da Documentação e Informação, Vila do Conde, 27 abril 2006*. Consultado en: junio 2008.

http://www.nosolousabilidad.com/hassan/visualizacion_y_recuperacion_de_informacion.pdf

He, Qin. "Knowledge discovery through co-word analysis". *Library trends*, 1999, v. 48, n. 1, pp. 133-159.

Herrero-Solana, Víctor; Hassan-Montero, Yusef. "Metodologías para el desarrollo de interfaces visuales de recuperación de información: análisis y comparación". *Information research: an international electronic journal*, 2006, v. 11, n. 3. Consultado en: abril 2008.
<http://eprints.rclis.org/archive/00006616/02/paper258.pdf>

Hildreth, Charles R. "The online acces catalog". *Annual review of information science and technology*, 1985, v. 20, pp. 233-285.

Janecek, Paul; Pu, Pear. "An evaluation of semantic fisheye views for opportunistic search in an annotated image collection". *International journal of digital libraries*, 2005, v. 5, n. 1, pp. 42-56.

Játiva-Miralles, M. Victoria. "Indicadores de calidad aplicables al análisis, evaluación y comparación de opacs". *El profesional de la información*, 2004, v. 13, n. 1, pp. 28-46.

Kamada, Tomihisa; Kawai, Satoru. "An algorithm for drawing general undirected graphs". *Information processing letters*, 1989, n. 31, pp. 7-15.

Korn, Flip; Shneiderman, Ben. *MeSH browse: A tool for browsing medical terms*, 1997. Consultado en: setiembre 2008.
<http://hcil.cs.umd.edu/trs/96-01/96-01.html>

Larson, Ray R. "Classification clustering, probabilistic information retrieval and the online catalog". *Library quarterly*, 1991, v. 61, n. 2, pp. 133-173.

Latour, Bruno. "Where are the missing masses? The sociology of a few mundane artifacts". En: **Bijker, W. E.; Law, J.** *Shaping technology/building society: Studies in sociotechnical change*. Cambridge: MIT Press, 1992, pp. 151-180. Consultado en: junio 2008.
<http://www.bruno-latour.fr/articles/article/50-MISSING-MASSESrepub.pdf>

Lin, Xia; Soergel, Dagobert; Marchionini, Gary. "A Self-organizing semantic map for information retrieval". *Proceedings of the 14th Annual international ACM Sigir conference on research and development in information retrieval*, 1991, pp. 262-269.

Marcos-Mora, Mari-Carmen. "Algunos principios básicos del diseño de interfaces de usuario" En: **Rovira, C.; Codina, L.** (dir.). *Documentación digital*. Barcelona: Universidad Pompeu Fabra, 2003. Consultado en: septiembre 2008.
http://www.mcmarcos.com/pdf/2002_algunos-modd.pdf

Marcos-Mora, Mari-Carmen. "La visualización en el proceso de búsqueda y recuperación de información". En: **Rovira, C.; Codina, L.** (dir.). *Documentación digital*. Barcelona: Universidad Pompeu Fabra, 2004. Consultado en: junio 2008.
http://www.mcmarcos.com/pdf/2004_visualizacion-modd.pdf

Marcos-Mora, Mari-Carmen; Baeza-Yates, Ricardo; Ardila, Carlos-Andrés. *Chile: A visual library catalog retrieval prototype*. Amsterdam: Springer, 2005. Consultado en: octubre 2008.
http://www.semanticweb.net/archives/2005_chile-springer.pdf

Matthews, Joseph R. *Public access to online catalogs*. New York: Neal-Schuman, 1983.

Mi, Jia; Weng, Cathy. "Revitalizing the library opac: Interface, searching, and display challenges". *Information technology & libraries*, 2008, v. 27, n. 1, pp. 5-22.

Miceli, Jorge E. "Los problemas de validez en el análisis de redes sociales: Algunas reflexiones integradoras". *Redes: Revista hispana para el análisis de redes sociales*, 2008, v. 14, n. 1. Consultado en: agosto 2008.
http://revista-redes.rediris.es/html-vol14/vol14_1.htm

Morillo-Velarde, José. "Aquadbrowser en Ceunet: implantación de un sistema de recuperación de información adaptado a las necesidades de una red de bibliotecas heterogéneas". En: *Proceedings VIII Jornadas de gestión de la información*, 2006, pp. 169-180, Madrid, España. Consultado en: septiembre 2008.
http://www.sedic.es/JornadasGI06_Jose_Morillo.pdf

Moya-Anegón, Félix de; Herrero-Solana, Víctor. "Investigaciones en curso sobre interfaces gráficos en dos y tres dimensiones para el acceso a la

información electrónica". *Cuadernos de documentación multimedia*, 1999, n. 8. Consultado en septiembre 2008.

<http://www.ucm.es/info/multidoc/multidoc/revista/num8/moya.html>

Moya-Anegón, Félix; Vargas-Quesada, Benjamín; Herrero-Solana, Víctor; Chinchilla-Rodríguez, Zaida; Corera-Álvarez, Elena; Muñoz-Fernández, Francisco-José. "A new technique for building maps of large scientific domains based on the cocitation of classes and categories". *Scientometrics*, 2004, v. 61, n. 1, pp. 129-145.

Noyons, Ed C. M.; Moed, Henk F.; Luwel, Mark. "Combining mapping and citation analysis for evaluative bibliometric purposes: a bibliometric study". *Journal of the American Society for Information Science*, 1999, v. 50, n. 2, pp. 115-131.

Peat, Helen J.; Willett, Peter. "The limitations of term co-occurrence data for query expansion in document retrieval systems". *Journal of the American Society for Information Science*, 1991, v. 42, n. 5, pp. 378-383.

Pichinini, Mariana; Martínez, Ana-María. "La búsqueda por materia en el catálogo en línea: los problemas del usuario". *Boletín bibliotecológico de La Plata*, 1996, v. 4, pp. 1-12.

Ríos-García, Yolanda. "Catálogos en línea de acceso público: selección bibliográfica". *Revista española de documentación científica*, 1991, v. 14, n. 2, pp. 121-141.

Rodríguez, Joseph A. "Análisis estructural y de redes". *Cuadernos metodológicos*, 1995, n. 16.

Schneiderman, Ben. "The eyes have it: A task by data type taxonomy for information visualization". En: *Proceedings of IEEE Symposium on visual languages*, 1996, pp. 336-343.

Senso, José A. "Visualizar información: ¿llegó la hora de las bibliotecas?". *El profesional de la información*, 2006, v. 15, n. 4, pp. 248-250. Consultado en: junio 2008.
<http://www.elprofesionaldeinformacion.com/contenidos/2006/julio/1.pdf>

Su, Shiao-Feng. "Dialogue with an opac: how visionary was Swanson in 1964?" *Library quarterly*, 1994, v. 64, n. 2, pp. 130-61.

Vargas-Quesada, Benjamín. *Visualización y análisis de grandes dominios científicos mediante redes pathfinder (Pfnnet)*. Granada: Universidad de Granada, 2005. Tesis doctoral.

Vargas-Quesada, Benjamín; Moya-Anegón, Félix de. *Visualizing the structure of science*. Berlin: Springer, 2007.

Villén-Rueda, Luis. "Indización y recuperación por materias en los opacs de las bibliotecas españolas: ¿dos décadas de evaluación?". *El profesional de la información*, 2006, v. 15, n. 2, pp. 87-98.

Wasserman, Stanley; Faust, Katherine. *Social network analysis: methods and applications*. Cambridge: Cambridge University Press, 1998.

White, Howard D.; McCain, Katherine W. "Visualization of literature". *Annual review of information science and technology*, 1997, v. 32, pp. 99-168.

White, Howard D. "Pathfinder networks and author cocitation analysis: A re-mapping of paradigmatic information scientists". *Journal of the American Society for Information Science and Technology*, 2003, v. 54, n. 5, pp. 423-434.

Sandra Miguel, Departamento de Bibliotecología, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata, 48 e/ 6 y 7, 5º piso. 1900 La Plata, Buenos Aires, Argentina. sandra@fcnym.unlp.edu.ar

Lorena Caprile e Israel Jorquera-Vidal, Biblioteca, Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Paseo del Bosque, s/n. 1900 La Plata, Buenos Aires, Argentina. lorenacaprile@yahoo.com.ar jorqueravidal@gmail.com



baratz

gestionando el conocimiento

Soluciones integrales para la gestión de:

- Bibliotecas
- Archivos
- Contenidos empresariales

Tel: 34 91 456 0360

informa@baratz.es

www.baratz.es